

Connecting the time domain community with the Virtual Astronomical Observatory

Matthew J. Graham^a, S.G. Djorgovski^a, Ciro Donalek^a, Andrew J. Drake^a, Ashish A. Mahabal^a, Raymond L. Plante^b, Jeffrey Kantor^c and John C. Good^d

^aCalifornia Institute of Technology, 1200 E California Blvd, Pasadena, CA 91125, USA;

^bNational Center for Supercomputing Applications, University of Illinois Urbana-Champaign, 1205 W. Clark St, Urbana, IL 61801, USA; ^c LSST Corporation, 950 N Cherry Ave, M-18, Tucson, AZ 85719, USA; ^d Infrared Processing and Analysis Center, California Institute of Technology, 770 South Wilson Ave., Pasadena, CA 91125, USA

ABSTRACT

The time domain has been identified as one of the most important areas of astronomical research for the next decade. The Virtual Observatory is in the vanguard with dedicated tools and services that enable and facilitate the discovery, dissemination and analysis of time domain data. These range in scope from rapid notifications of time-critical astronomical transients to annotating long-term variables with the latest modelling results. In this paper, we will review the prior art in these areas and focus on the capabilities that the VAO is bringing to bear in support of time domain science. In particular, we will focus on the issues involved with the heterogeneous collections of (ancillary) data associated with astronomical transients, and the time series characterization and classification tools required by the next generation of sky surveys, such as LSST and SKA.

Keywords: Time domain, Virtual Observatory, data access, characterization, classification

1. INTRODUCTION

The time domain is the emerging field of astronomical research, as recognized in the 2010 National Research Councils Decadal Survey of Astronomy and Astrophysics.¹ Planned facilities for the next decade and beyond, such as the Large Synoptic Survey Telescope (LSST)* and the Square Kilometer Array† (SKA), will revolutionize our understanding of the universe with nightly searches of large swathes of sky for changing objects and networks of robotic telescopes ready to follow up in greater detail selected interesting sources. This will impact essentially every area of astronomy, from the Solar System to cosmology, and from stellar evolution to extreme relativistic phenomena,² making it a very rich area for scientific exploration and discovery. Moreover, many interesting phenomena, e.g., supernovae and other types of cosmic explosions, can only be studied in the time domain.

These new surveys build on a legacy of over fifty years of experience of sky surveys, first with photographic plates and then, more recently, digital detectors (see Ref. 3 for a recent review). The rise of information technology has driven an exponential growth of data volumes (and, equally importantly, data complexity and data quality) following Moore's law, e.g., DPOSS⁴ to 2MASS⁵ to SDSS,⁶ and many digital sky surveys that followed. To cope with such (necessarily distributed) giga- and terascale data collections, the community developed the concept of the *Virtual Observatory* (VO), which provides the wherewithal to aggregate and analyze disparate data sets, opening up new avenues of scientific research based on data discovery and fusion.^{7,8} The Virtual Astronomical Observatory⁹ (VAO) is the US national VO project and provides the components, libraries, and templates that allow national facilities, major projects, and end-users to craft their own VO-enabled applications for seamless data access and integration, especially in support of data intensive research.

Further author information: (Send correspondence to M.J.Graham)

M.J.G.: E-mail: mjg@caltech.edu, Telephone: 1 626 395 8030

*<http://www.lsst.org>

†<http://www.skatelescope.org>

Table 1. Different types of data access protocol defined by the IVOA.

Name	Description
Simple Cone Search (SCS)	Retrieve all objects within a circular region on the sky
Simple Image Access (SIA)	Retrieve all images of objects within a region on the sky
Simple Spectral Access (SSA)	Retrieve all spectra of objects within a region on the sky
Simple Line Access (SLA)	Retrieve spectral line data
Simulations (SIMDAL)	Retrieve simulation data
Table Access (TAP)	Retrieve tabular data

The time domain has been identified early as a prime arena for VO applications.¹⁰ It adds a new dimension to data discovery and federation with (near) real-time massive data streams - for example, Palomar-Quest,¹¹ Catalina Real-time Transient Survey¹² (CRTS), Palomar Transient Factory¹³ (PTF) and Panoramic Survey Telescope And Rapid Response System¹⁴ (Pan-STARRS; PS1) and LSST to come - replacing static data sets; in a way, we've moved from panoramic digital photography of the sky to panoramic digital cinematography. Since many of the observed phenomena in this domain are short-lived, and since the scientific returns depend strongly not only on their detection, but also on the timely and well-chosen follow-up observations, there is a need to fully process the data as they stream from the telescopes, compare it with the previous images of the same parts of the sky, automatically and reliably detect any changes, and classify and prioritize the detected events for the rapid follow-up observations. This poses significant new technological challenges for the VO and its infrastructure. Analogous situations may also be found in many other areas, where the data come continuously from some instruments or sensor networks, and where anomalous or specifically targeted events have to be found and responded to in a rapid fashion.

The VO has evolved a two-track approach to the time domain: one deals specifically with the mechanics of reporting transient celestial events (VOEvent) in a timely fashion and the associated infrastructure to publish, disseminate and archive them. The other deals with the more general issues of time series data, such as how to describe, represent and access them in a way to ensure interoperability between different data archives. In this paper, we will review the specific issues that are associated with the latter and how the VO, and, specifically, the VAO, which is leading the time domain effort, is meeting them. This draws in associated but separate work on source characterization and classification that is an essential part of a time domain system. Details of the transient approach are presented in a complementary paper in these proceedings,¹⁵ although we will discuss certain common issues. Nevertheless, whichever approach is being discussed, operational concerns are an important consideration and we focus particularly on those related to questions of scalability and managing large collections of heterogeneous data.

2. INTEROPERABLE TIME SERIES

The promise of data federation is that it can often lead to potentially new scientific insights. An obvious example is combining observations of the same objects from different wavelength regimes, e.g., X-ray, infrared, and radio, to understand the various physical processes that contribute to their spectral energy distributions. The time domain adds an extra dimension to this, allowing the identification of temporally correlated behavior, for example, a X-ray burst followed by an infrared burst may indicate the propagation of a shock front from an originating source to circumscribing material.

The International Virtual Observatory Alliance[‡] (IVOA) has defined a common set of data access protocols to ensure that the same interface is employed across all data archives, no matter where they are located, to perform the same type of data query (see Table 1 for a summary of those defined). Although common data formats may be employed in transferring data, e.g., VOTable¹⁶ for tabular data, individual data providers usually represent and store their data and metadata in their own way. Common data models define the shared elements across data and metadata collections and provide a framework for describing relationships between them so that different representations can interoperate in a transparent manner.

[‡]<http://www.ivoa.net>

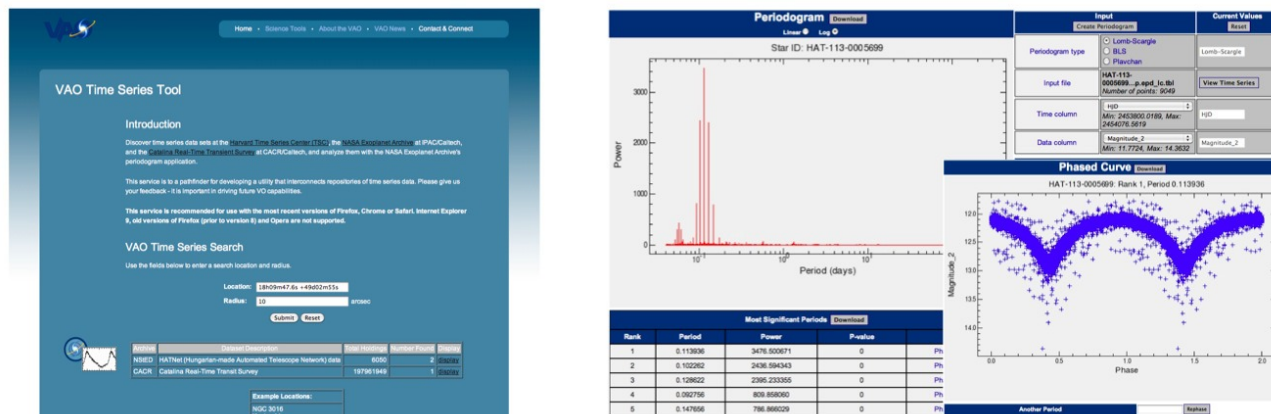


Figure 1. This shows a screenshot from the VAO Time Series Search Tool (left) and the Lomb-Scargle periodogram (right) and phased light curve (inset).

VOEvent¹⁷ may be regarded as the (lightweight) data model for observations of transient astronomical events, describing their who, what, where/when, how, and why characteristics. However, when individual measurements of arbitrarily named quantities are reported, either as a group of parameters or in a table, their broader context within a standard data model can be established through the IVOA Utypes mechanism.¹⁸ These strings act as reference pointers to individual elements within a data model thus identifying the concept that the reported value represents, e.g., the UType “Data.FluxAxis.Accuracy.StatErrHigh” identifies a quantity as the upper error bound on a flux value defined in the Spectral data model. Namespaces allow reuse of quantities/concepts defined in one data model in another.

The Spectral data model¹⁹ (now in the final stages of specification) defines a generalized model for spectrophotometric sequences and provides a basis for a set of specific case models, such as Spectrum, SED and TimeSeries. The TimeSeries data model is intended to describe any observed or derived quantity that may vary with time with a light curve being the most common example and considered to be a time series with just one photometric band. More complicated time series might involve multi-band data or data where the time sample bin size varies between successive samples. The data model can also describe various levels of associated metadata such as period for the time series if relevant or known and whether the time axis values are folded by the period if one exists, or a target variability amplitude and derived signal-to-noise ratio. A variety of serialization formats for compliant data sets are supported, e.g., FITS, VOTable, and CSV.

A number of data archives are set to expose their time series holdings via the TimeSeries data model. Prime amongst these is the Catalina Real-time Transient Survey¹² (CRTS) which has light curves for several hundred million objects over ~ 33000 deg² between $-75^\circ < Dec < 75^\circ$ (except for within $\sim 10 - 15^\circ$ of the Galactic plane) to ~ 20 mag. and with an average of ~ 250 observations over a 7-year baseline. ~ 200 million light curves from the CRTS DR1 are already accessible from both the CRTS web site[§] and the VAO Time Series Search Tool[¶]. The latter site (see Fig. 1) is a pathfinder utility for interconnecting repositories of time series data, and provides access to important data sets at the NASA Exoplanet Archive^{||}, such as Kepler, CoRoT, HATNet, TrES and KELT, and at the Harvard Time Series Center^{**}, such as ASAS, in addition to the CRTS DR1 data. The service also offers access to some analysis tools (see below).

3. EVENT AND SOURCE CROSS-IDENTIFICATION

A component of the VOEvent architecture¹⁷ is an *annotator*, which is a service that receives an event notification and acts on it in such a way that a subsequent event is generated adding information to what is already known

[§]<http://nesssi.cacr.caltech.edu/DataRelease>

[¶]<http://www.usvao.org/science-tools-services/time-series-search-tool>

^{||}<http://exoplanetarchive.ipac.caltech.edu>

^{**}<http://timemachine.iic.harvard.edu/>

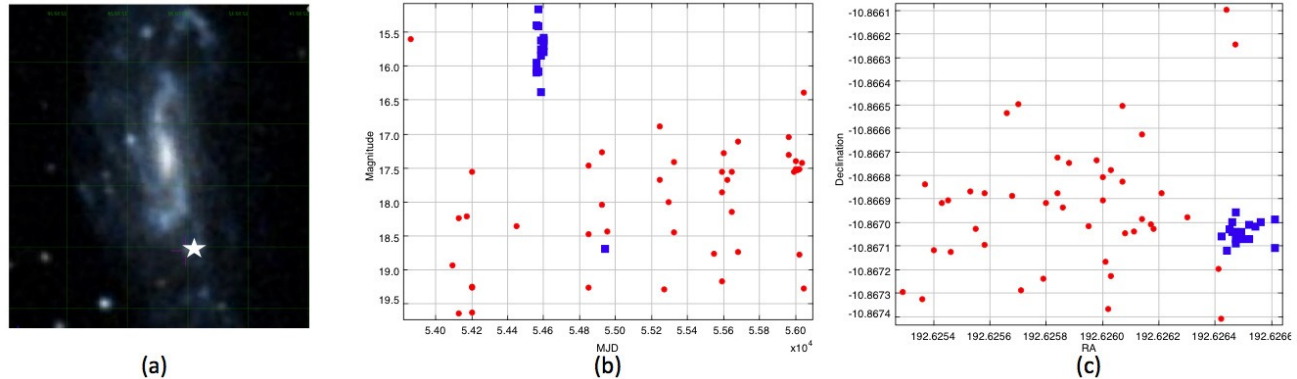


Figure 2. This shows how positional cross-matching information can be used to correctly identify variable sources. The full signal from the supernova 2008aq (a) in the CRTS light curve (b) (squares) only becomes apparent when the spatial positions are examined (c).

about the celestial event that triggered the original event notification. One of the most common annotation activities is to cross-identify events with other data archives, i.e., search for plausible spatial associations between an event and other observations, typically at other wavelengths (the VAO Cross-Comparison Tool^{††} performs fast positional cross-matches between an input table of up to 1 million sources and common astronomical source catalogs, such as SDSS and 2MASS). The simplest matching criterion is just to take the nearest positional hit but this is not necessarily the best match. Positional accuracies can vary widely between surveys, particularly between different wavelength regimes, leading to multiple possible crossmatch candidates, e.g., a brighter object might have a smaller positional error due to its stronger detection whereas a fainter object might be farther away yet still as likely due to its larger positional error. Other information may also make certain matches far more likely, such as a potential supernova being more likely associated with a nearby galaxy than a star. Several formalisms have been proposed to deal with the general problem of spatial crossmatching but Budavari²⁰ uses Bayesian hypothesis testing to evaluate the quality of candidate associations specifically for detections in space *and* time thus allowing inclusion of information about the temporal behavior of particular sources.

A related activity is constructing the time histories of astronomical objects from sets of individual observations of them within the same survey. Fig. 2(b) shows the CRTS time series for the position associated with SN 2008aq. The light curve for the supernova (squares) is contaminated by observations of nearby sources (points) since it occurred in the outer reaches of a spiral galaxy (MCG-02-33-20, Fig. 2(a)). Only by crossmatching the positions of individual observations can the subgroup that corresponds to the actual supernova be identified; in particular, the late data point from the supernova when it has faded by ~ 3 mags. which would otherwise be lost in the background signal from the galaxy. The Budavari formalism can recover this information on the assumption of a suitable prior model for the supernova light curve.

Recovering the light curve for a single source is generally a straightforward operation but for every object in a survey, it soon becomes intractable. Constructing the (full transitive) set of associations for n sources from a set of m observations scales at best as $\mathcal{O}(\sim nm^2)$, assuming only one match between individual sets (note that $\mathcal{O}(\sim n^2)$ is normally already considered prohibitive in many circles). The Palomar-Quest DR1 (PQDR1) catalog of ~ 10 million sources with typically ~ 15 observations per source resulted in a set of over 4 billion associations; with ~ 500 million sources, each with typically ~ 200 observations per source, CRTS would have at least ~ 20 trillion associations. The next generation surveys will take us into the quadrillions and beyond. These are also inherently probabilistic groupings. Varying conditions between observations – sky brightness, atmospheric, instrumental, etc. – mean that the same detection thresholds and positional errors cannot be assumed across a survey, e.g., perfect conditions may mean that two nearby sources are resolved on one night but appear as a blended source on another poorer quality night so a single observation may well be associated with multiple histories, particularly in crowded fields.

^{††}<http://www.usvao.org/science-tools-services/cross-comparison-tool>

An obvious scalable solution to maintain a master catalog and update associations on a per night basis (using a service such as the VAO Cross-Comparison Tool) rather than associate all data in a single go. However, this latter operation will still become necessary when the master catalog needs to be revised, e.g., with improved positional error models. It is interesting to note that whilst spatial pixellation schemes, which provide a single identifier for a region of sky, such as HTM²¹ and Healpix,²² are good for individual object lookups in catalogs, there are more efficient ways of doing bulk crossmatches. The Zones algorithm²³ developed for the SDSS and 2MASS surveys uses a B-tree to bucket two-dimensional space giving dynamically computed bounding boxes (B-tree ranges) for spatial queries. In practice, using an optimal zoning gives several factors of ten increased performance over using indexing schemes, although, in tests with PQDR1, the Quad Tree Cube scheme²⁴ has also shown itself to be equivalently fast.

4. SOURCE CHARACTERIZATION

When a significant variation in an astronomical source is detected (the significance is determined by such factors as the size, suddenness and duration of the variation as well as the type of detector used), an event notification (VOEvent) is broadcast to all interested parties. This triggers a cascade of activity where the event is placed in context with related data and information: followup observations of the same astronomical event (if possible), source cross-identifications, etc. This *data portfolio* for the event represents a summation of all that is known and understood about it. The most interesting or exciting events will be associated with rich portfolios containing a wide range of heterogeneous material whereas a commonplace event might have a portfolio containing only the original event notification. Note that a portfolio is also a dynamic entity with the potential for new material to be added at any time, from milliseconds to years or even decades after the initial event.

For any analysis activity involving the comparison of events (or sources), the heterogeneity of information has to be replaced by a common characterization in terms of a representative set of features. Even when dealing just with light curves, there can be tremendous disparity between temporal coverage, sampling rates and regularity, number of points and error bars, e.g., a transient light curve might consist of a couple of points and a lot of upper limits (when the source was below the detection threshold of the survey) whilst a monitored exoplanet candidate light curve may contain hundred of thousands of high signal-to-noise points. A number of recent papers^{25–27} have explored a variety of characterizing features for the light curves of known variable stars, including statistical moments, flux and shape ratios, variability indices, periodicity measures, and model representations. In this vein, the Caltech Time Series Characterization Service^{††} (CTSCS) is an experimental service which aims to extract a comprehensive set of (over 60) such features from any supplied light curve.

Many features employed in characterizing light curves are founded on some type of periodic analysis of the time history of the object. Unfortunately there is no single way to determine the period of a light curve that is accurate and reliable - for example, the Lomb-Scargle method, which is in many ways the de facto standard technique, is at best ~75% accurate and less so with irregular sampling strategies and light curves with small numbers of points.²⁸ The NASA Exoplanet Archive Periodogram Service* supports three algorithms: Lomb-Scargle, box-fitting least squares and Plavchan, as well as control over their various parameters so that they can be fine-tuned for a particular type of time series data. An advantage of this service is that it is integrated with the VAO Time Series Search Tool so a period can be easily obtained for a discovered light curve without having to download the data first and then upload it to the service. The CTSCS offers more period-finding algorithms but is not integrated with the Time Series Tool. The NASA tool is also designed around an enterprise architecture and so can handle larger workloads.

5. SOURCE CLASSIFICATION

Source classification is a very challenging problem, particularly for transient events, due to the sparsity and heterogeneity of the available data. Some current efforts on the classification of transients in the optical synoptic sky surveys, employing both the source light curve and its data portfolio, include Refs. 3,29–34. The Caltech-JPL

^{††}<http://nirgun.caltech.edu:8000>

*<http://exoplanetarchive.ipac.caltech.edu/cgi-bin/Periodogram/nph-simpleupload>

group, in particular, has experimented with several approaches, using data from the Palomar-Quest survey¹¹ and Catalina Real-time Transient Survey,^{12,34,35} as well as selected data sets from the literature.

When addressing archival (i.e., not real time) classification, light curves are reduced to common basis feature vectors (as described above) as input to machine learning techniques (e.g., decision trees (DT)). In particular, DTs have been trained using the feature vectors for various combinations of known classes of object. To reduce the dimensionality of the input space, a forward feature selection strategy is applied that selects the best subset of features from a training data set that can predict test data. The DTs themselves are built using the Gini diversity index as the splitting criterion with 10-fold cross validation to avoid overfitting. Moreover, the DTs in the iterations are pruned in order to choose the simplest one within one standard error of the minimum. When tested on a data set consisting of light curves of blazars, cataclysmic variables, and RR Lyrae stars, the method³⁶ achieves between $\sim 83\%$ and 97% completeness and $\sim 4\%$ to 13% contamination. This approach seems very promising for the classification of radio light curves as well.

Another novel approach^{37,38} uses 2-dimensional distributions of magnitude changes for different time baselines for all possible epoch pairs in the data set. These 2-dimensional $(\Delta m, \Delta t)$ histograms can be viewed as probabilistic structure functions for the light curves of different types. Template distributions for different kinds of transients and variables are constructed using the reliably-classified data with the same survey cadences, S/N, etc. For any newly detected variable or a transient, corresponding $(\Delta m, \Delta t)$ histograms are accumulated as the new data arrive, and a variety of metrics used to compute the effective probabilistic distances from different templates. The tests so far indicate that classification accuracies in excess of 90% may be possible using this approach. Generalizations to include triplets or even higher order sets of data points for multi-dimensional histograms are planned.

Other approaches have been tested as well but their description is beyond the scope of this paper. However, it quickly became very clear that different classifiers may be optimal for different types of transient or variable sources. Thus, there is a need for a meta-classifier that would provide an optimal classification for any given event on the basis of several different classifier results. This work is still in progress.

One important lesson learnt so far is that the existing archival and contextual data will play a critical role in classification. This includes a spatial context (i.e., what is near the observed event on the sky - a possible host galaxy, a cluster, a SN remnant, etc.), the multiwavelength context (has it been detected on other wavelength, what is its broad spectral energy distribution), and the temporal context (what was its flux variability or a detection history in the previously obtained data). In fact, most of the relevant information in hand when a transient is first detected is of this nature. Some of it can be readily extracted from the archives using VAO tools (e.g., flux measurements from different wavelengths, light curves at that location), but some - spatial context in particular - require a human judgment, e.g., is the apparent proximity to other objects (galaxies, clusters, etc.) likely to be relevant, and if so, what does it imply about the transient? Human inspection of vast numbers of transients does not scale to the massive data streams such as those contemplated here. Crowdsourcing approaches to harvesting the relevant human pattern recognition skills and domain expertise, and their translation to machine-processable algorithms are being investigated.

It is intended that automated transient source classifiers for CRTS detections will be deployed as annotator services in the near future. Classifications for individual events will be broadcast as followup VOEvents and employ a machine-processable concept scheme, e.g., the IVOA thesaurus or Unified Astronomy Thesaurus, to describe the most likely object type. This will allow robotic telescopes to trigger followup observations not only on parameter-based rules, e.g., magnitude < 17 , but also on (probabilistic) classifications described with community-standard terms.

6. CONCLUSIONS

The emerging field of time domain astronomy requires tools and infrastructure to support a distributed network of (massive) real-time data streams, data archives, and analysis services. The VAO is developing an interoperable framework to connect partner providers of both data and analysis resources, and expose them as an integrated whole for wider community use. A recent community-wide call for collaborative proposals by the VAO⁹ has identified two time domain projects which it is now advising. One is concerned with access to data related to

the Variable and Slow Transient (VAST) Survey Science Program of the Australian Square Kilometre Array Pathfinder (PI: T. Murphy) and the other involves access to the databases of the American Association of Variable Star Observers (AAVSO, PI: M. Templeton). Such collaborations, combining domain expertise in data technologies and the relevant science areas, illustrate the potential of an informatics-based approach to data-intensive science.

ACKNOWLEDGMENTS

Support for the development of time series infrastructure is provided by the Virtual Astronomical Observatory contract AST-0834235. Work on source characterization and classification has been supported in part by the National Science Foundation grants AST-0407448, CNS-0540369, AST-0909182 and IIS-1118041; the National Aeronautics and Space Administration grant 08-AISR08-0085; and by the Ajax and Fishbein Family Foundations. We are thankful to numerous colleagues in the VO and Astrominformatics community, and to the members of the DPOSS, PQ, and CRTS survey teams, for many useful discussions and interactions through the years.

REFERENCES

- [1] Survey of Astronomy and Astrophysics Committee, National Research Council, *New Worlds, New Horizons in Astronomy and Astrophysics*, The National Academies Press, New York, 2010.
- [2] S. G. Djorgovski, A. Mahabal, A. Drake, M. Graham, C. Donalek, and R. Williams, “Exploring the time domain with synoptic sky surveys,” in *New Horizons in Time-Domain Astronomy*, R. Griffin, R. Hanisch, and R. Seaman, eds., *IAU Symp.* **285**, pp. 141–147, 2012.
- [3] S. G. Djorgovski, A. A. Mahabal, A. J. Drake, M. J. Graham, and C. Donalek, “Sky surveys,” in *Planets, Stars, and Stellar Systems*, T. Oswald, ed., Springer Verlag, Berlin, 2012.
- [4] S. G. Djorgovski, R. Gal, S. Odewahn, R. de Carvalho, R. Brunner, G. Longo, and R. Scaramella, “The Palomar Digital Sky Survey (DPOSS),” in *Wide Field Surveys in Cosmology*, S. Colombi, Y. Mellier, and B. Raban, eds., p. 89, 1998.
- [5] M. Skrutskie and the 2MASS team, “The Two Micron All Sky Survey (2MASS),” *AJ* **131**, p. 1163, 2006.
- [6] D. G. York and the SDSS team, “The Sloan Digital Sky Survey: Technical Summary,” *AJ* **120**, p. 1579, 2000.
- [7] R. Brunner and S. G. Djorgovski in *Virtual Observatories of the Future*, R. J. Brunner, S. G. Djorgovski, and A. S. Szalay, eds., *Astronomical Society of the Pacific Conference Series* **225**, p. 52, 2001.
- [8] S. G. Djorgovski and R. Williams, “Virtual Observatory: From Concept to Implementation,” in *From Clark Lake to the Long Wavelength Array: Bill Erickson’s Radio Science*, N. Kassim, M. Perez, W. Junor, and P. Henning, eds., *Astronomical Society of the Pacific Conference Series* **345**, p. 517, 2005.
- [9] G. B. Berriman, R. J. Hanisch, J. W. Lazio, A. Szalay, and G. Fabbiano, “The organization and management of the Virtual Astronomical Observatory,” in *Modeling, systems engineering, and project management for astronomy V*, G. Z. Angeli and P. Dierickx, eds., *Proc. SPIE* **8449**, 2012.
- [10] S. G. Djorgovski et al., “Exploration of Large Digital Sky Surveys,” in *Mining the Sky*, A. J. Banday, S. Zaroubi, and M. Bartelmann, eds., p. 305, 2001.
- [11] S. G. Djorgovski et al., “The Palomar-Quest digital synoptic sky survey,” *AN* **329**, p. 263, 2008.
- [12] A. J. Drake and the CRTS team, “First results from the Catalina Real-time Transient Survey,” *ApJ* **696**, p. 870, 2009.
- [13] A. Rau et al., “Exploring the Optical Transient Sky with the Palomar Transient Factory,” *PASP* **121**, pp. 1334–1351, 2009.
- [14] N. Kaiser et al., “Pan-STARRS: A Large Synoptic Survey Telescope Array,” *SPIE Conference Series* **4836**, pp. 154–164, 2002.
- [15] R. D. Williams, S. D. Barthelmy, R. B. Denny, M. J. Graham, and J. Swinbank, “Responding to the event deluge,” in *Observatory Operations: Strategies, Processes and System IV*, A. B. Peck, R. L. Seaman, and F. Comeron, eds., *Proc. SPIE* **5884**, 2012.
- [16] F. Ochsenbein et al., “VOTable Format Definition Version 1.2,” *IVOA Recommendation*, 2011.
- [17] R. Seaman et al., “Sky Event Reporting Metadata Version 2.0,” *IVOA Recommendation*, 2011.

- [18] M. Louys et al., “Utypes: a standard for serializing data models instances,” *IVOA Working Draft*, 2012.
- [19] M. Cresitello-Dittmar et al., “IVOA Spectral Data Model,” *IVOA Working Draft*, 2012.
- [20] T. Budavari, “Probabilistic cross-identification of cosmic events,” *ApJ* **736**, pp. 155–160, 2011.
- [21] P. Z. Kunszt, A. S. Szalay, and A. R. Thakar, “The hierarchical triangular mesh,” in *Mining the Sky*, p. 631, 2001.
- [22] K. M. Gorski et al., “HEALPix: A Framework for High-Resolution Discretization and Fast Analysis of Data Distributed on the Sphere,” *ApJ* **622**, p. 759, 2005.
- [23] J. Gray, M. Nieto-Santisteban, and A. S. Szalay, “The Zones Algorithm for Finding Points-Near-a-Point or Cross-Matching Spatial Datasets,” 2007.
- [24] S. Kposov and O. Bartunov, “Q3C, Quad Tree Cube - The new Sky-indexing Concept for Huge Astronomical Catalogues and its Realization for Main Astronomical Queries (Cone Search and Xmatch) in Open Source Database PostgreSQL,” in *Astronomical Data Analysis Software and Systems XV*, C. Gabriel, C. Arviset, D. Ponz, and E. Solano, eds., p. 735, 2006.
- [25] J. Richards et al., “On Machine-learned Classification of Variable Stars with Sparse and Noisy Time-Series Data,” *ApJ* **733**, p. 10, 2011.
- [26] P. Dubath et al., “Random forest automated supervised classification of Hipparcos periodic variable stars,” *MNRAS* **414**, pp. 2602–2617, 2011.
- [27] M.-S. Shin, M. Sekora, and Y.-I. Byun, “Detecting variability in massive astronomical time series data - I. Application of an infinite Gaussian mixture model,” *MNRAS* **400**, pp. 1897–1910, 2009.
- [28] M. J. Graham, A. J. Drake, S. G. Djorgovski, A. A. Mahabal, and C. Donalek, “A comparison of period-finding algorithms,” 2012, in prep.
- [29] J. S. Bloom, D. L. Starr, N. R. Butler, P. Nugent, M. Rischard, D. Eads, and D. Poznanski, “Towards a real-time transient classification engine,” *AN* **329**, p. 284, 2008.
- [30] C. Donalek et al., “New Approaches to Object Classification in Synoptic Sky Surveys,” in *American Institute of Physics Conference Series*, C. Bailer-Jones, ed., **1082**, pp. 252–256, 2008.
- [31] A. Mahabal et al., “Towards Real-time Classification of Astronomical Transients,” in *American Institute of Physics Conference Series*, C. Bailer-Jones, ed., **1082**, pp. 287–293, 2008.
- [32] A. Mahabal et al., “Automated probabilistic classification of transients and variables,” *AN* **329**, pp. 288–291, 2008.
- [33] A. Mahabal et al., “Mixing Bayesian Techniques for Effective Real-time Classification of Astronomical Transients,” in *Astronomical Data Analysis Software and Systems XIX*, **434**, p. 115, 2010.
- [34] A. Mahabal et al., “Discovery, classification, and scientific exploration of transient events from the Catalina Real-time Transient Survey,” *Bulletin of the Astronomical Society of India* **39**, pp. 397–408, 2011.
- [35] S. G. Djorgovski et al., “The Catalina Real-time Transient Survey (CRTS),” in *The First Year of MAXI: Monitoring Variable X-ray Sources*, T. Mihara and M. Serino, eds., **127**, p. 263, 2012.
- [36] C. Donalek et al. 2012, in prep.
- [37] S. G. Djorgovski et al., “Towards an Automated Classification of Transient Events in Synoptic Sky Surveys,” in *Proc. CIDU*, A. Srivasatva, N. Chawla, and A. Perera, eds., p. 171, 2011.
- [38] B. Moghaddam et al. 2012, in prep.